From my point of view

4 Apr 2019 by Ahmad Fathi Syed

Public Blog

A curious cat looks into data

My unnamed manager (although you can still find out in the directory if you are a curious cat) tasked me to do some graphical view of 10-years historical data last quarter. It has been an educational task, which I am grateful to have a chance of doing. Also, my eternal thanks to John, Amanda, and local characteristic's team here for helping me extracting, filtering and arranging the data which is in bulk form.

Usually, our understanding of data (at least commonly perceive) is a solid substance – a matter of fact. But that is not usually the case. Yes, some of it is objective or hard data eg. length of a vessel, but some of it is a matter of opinion which need to be determined by judgement (of course with what information we had in hand). For example, a vessel severely burnt in her cargo hold – needing repairs but still can continue voyage by her own propulsion, is it 'serious' or 'non-serious' category?

To address (or at least reduce) the uncertainty and produce consistency, analyst in a team must agree on a specific definition for each type of data. This should solve the problem, right? Not so. Sometimes new factor comes in, new technologies were born, or the determining fact is sketchy, so some old definition might not give justice to the data.

Tags

Add a tag

Related files

Drop files here or select files

Up to this point, you might say, *"well you guys need to sit down and agree on a new set of definition, what's the big deal".* Fair point (but there is still a 'but'). I'm an avid reader, currently reading a book on income inequality by economist Matthew P. Drennan, interestingly he said in the book, *"If you collect a data for some time, and change the methodology in the middle, the new set of data is not comparable to the old set"* (I'm paraphrasing – not his actual quote). So, if we change the definition we are likely to get a variation of result.

Definition of data is important, especially when we want to do a comparison. When we had two set of data, if there is a difference, the first question should be: What is the definition? Otherwise, we will fall into a trap and drew a wrong conclusion leading to a wrong decision.

By this time, if you are lost, then we are on the same page. My intention is not to come up with a solution (if I can solve it maybe I can get a Nobel Peace Prize?), as a learner and a curious cat swimming in the ocean of data, my point is - data is not as straightforward as you think. I still believe the best way to solve them is through discussion, in many discussions, I come to a realization that I am not the smartest person on earth! And of course, there is a bright side, if everything is the same and consistent the world will be dull and boring. Problems, uncertainty, and differences are things that made life interesting – so do data (I managed to end this piece with a positive note! Now, rejoice).

Before you go, I want to suggest to you one interesting read on this topic, a small book titled *How to Lie with Statistics* by Darrell Huff. It will make you one kilogram cleverer and giggles at the same time. Cheers!

Reply here...

Careers

Career pathing tool